

УДК 004.021

Н.А. Попова, М.А. Назаров, М.В. Власов  
**РЕШЕНИЕ ЗАДАЧИ РАСПОЗНАВАНИЯ ЛИЦ С  
ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ**  
*Пензенский государственный университет, Пенза, Россия*

*Целью данной статьи является обобщение полученного опыта разработки и реализации нейронной сети для распознавания лиц. В основу нейронной сети входят специальные алгоритмы машинного обучения. В качестве входных данных алгоритм получает изображение с лицом одного человека или лицами нескольких людей, после чего происходит поиск всех лиц в данном изображении с использованием гистограмм направленных градиентов, ее результатом является фрагменты изображения, где явно проглядываются базовые структуры лица или лиц. Для того, чтобы определить уникальные черты лица, необходимо учитывать разность угла наклона лица и степень его освещенности, для этого на выделенных фрагментах применяется алгоритм оценки ориентиров для поиска 68 особых точек, которые существуют на каждом лице, полученные точки дают возможность как можно лучше отцентрировать глаза и рот для более точного кодирования. Кодирование изображения включает в себя построение точной "карты лица" состоящей из 128 измерений. На основе полученных результатов, сверточная нейронная сеть, используя алгоритм линейного классификатора SVM, может определять соответствие между разными фотографиями.*

**Ключевые слова:** распознавание лиц, машинное обучение, гистограмма направленных градиентов, HOG, оценка ориентиров лица, аффинные преобразования, глубокое обучение, нейронная сеть.

Интерес к процессам узнавания и распознавания лиц, всегда был значительным, особенно в связи со все возрастающими практическими потребностями: системы охраны, верификация кредитных карточек, криминалистическая экспертиза, телеконференции и т.д. Несмотря на ясность того житейского факта, что человек хорошо идентифицирует лица людей, совсем не очевидно, как научить этому компьютер, в том числе как декодировать и хранить цифровые изображения лиц. Одним из лучших вариантов решения данной проблемы может быть использование нейронных сетей.

Для создания нейронной сети используются различные алгоритмы машинного обучения. Проведя эмпирическое исследование наиболее популярных алгоритмов, были выбраны наиболее оптимальные из них.

Для решения поставленных задач можно выделить следующие основные шаги:

**1 Поиск всех лиц на фотографии:** необходим для того, чтобы выделить область изображения, которая передается в дальнейшую

обработку. Для этой цели используется **гистограмма направленных градиентов** (Histogram of Oriented Gradients - HOG) [1].

Согласно этому алгоритму, изображение делается черно-белым, потому что для поиска лиц цветные данные не нужны. В цикле просматривается каждый пиксель и его соседние, с целью – выяснить, насколько темным является текущий пиксель по сравнению с окружающими. Затем добавляется стрелка, показывающая, в каком направлении изображение становится темнее. Выполнив эту процедуру для каждого отдельного пикселя изображения, каждый пиксель заменяется стрелкой. Эти стрелки называются *градиентами*, и они показывают направление от светлых пикселей к темным по всему изображению. Если использовать темные и светлые изображения одного человека, пиксели будут иметь различные показатели яркости, но при рассмотрении направления изменений яркости, получается одинаковая картинка независимо от яркости исходного изображения. Для сохранения ресурсов и избавления от избыточной информации, изображение делится на блоки размером 16x16 пикселей. Затем заменим этот квадрат на изображении стрелками, направленными туда же, куда и большинство.

В конечном итоге исходное изображение преобразуется так, что ясно проглядывается базовая структура лица. Чтобы найти лицо на HOG-изображении, необходимо найти часть изображения, которая наиболее похожа на известный рисунок HOG, полученный из множества других лиц в ходе обучения. Пример фрагмента, выделяемого на изображении представлен на Рисунке 1.



Рисунок 1 – Поиск лица на фото

**2 Расположение лица:** для того чтобы компьютеру было проще работать с повернутыми лицами, необходимо преобразовать каждое

изображение так, чтобы глаза и губы всегда находились на конкретном месте. Для решения этой задачи используется алгоритм **оценки ориентиров лица** (*face landmark estimation*) [2]. Основная идея состоит в том, что отмечаются 68 особых точек (называемых *ориентирами*), которые существуют на каждом лице – верхняя часть подбородка, внешняя точка каждого глаза, внутренняя точка каждой брови и т. д. Затем обучается алгоритм машинного обучения чтобы находить эти 68 особых точек на любом лице. На Рисунке 2, 3 показаны выделенные ориентиры лица.

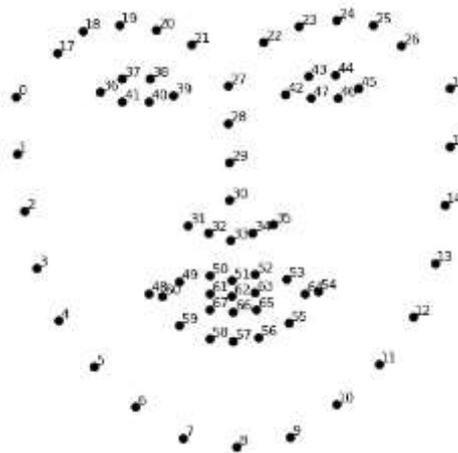


Рисунок 2 – 64 ориентира лица

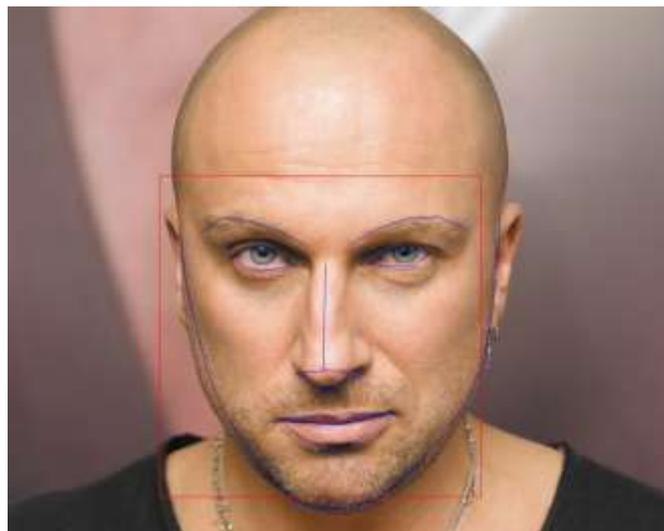


Рисунок 3 – Определение ориентиров лица

Когда на изображении определяются глаза и рот, его можно вращать, масштабировать и сдвигать так, чтобы глаза и рот были как можно лучше центрированы. Для центрирования используются только базовые преобразования изображений, такие как поворот и масштабирование, которые сохраняют параллельные линии - аффинные преобразования.

В основе этого метода преобразования изображения лежит следующий принцип: для каждой точки конечного изображения берется фиксированный набор точек исходного и интерполируется в соответствии с их взаимным положением. Аффинное преобразование является самым общим взаимно однозначным отображением плоскости на плоскость, при котором сохраняются прямые линии и отношения длин отрезков, лежащих на одной прямой. После этого преобразования лицо центрируется примерно в одно и то же положение на изображении, что делает следующий этап более точным.

**3 Кодирование лица:** полученное после первых шагов изображение лица необходимо “закодировать” уникальным образом, так как сравнивать полученное изображение со всеми предыдущими это нерациональный подход при наличии большой базы изображений. Значит нам нужен способ снять несколько основных измерений с каждого лица, которые мы могли бы сравнить с самыми близкими известными измерениями и найти самое похожее лицо. В 1960 году Вудроу Бледсоу предложил алгоритм определяющий очевидные черты лица. Однако эти измерения для мозга человека (цвет глаз, размер носа и тд), на самом деле не имеют смысла для компьютера, рассматривающего отдельные пиксели на изображении. Исследователи показали, что самый точный подход – это позволить компьютеру самому измерить то, что ему нужно. Глубокое обучение, определяет, какие части лица нужно измерять, лучше, чем люди [3].

Для решения этой проблемы создается сверточная нейронная сеть глубокого обучения, которая генерирует 128 измерений (карта лица) для каждого человека (Рисунок 4). Идея преобразования изображений, в список генерируемых компьютером чисел крайне важна для машинного обучения (особенно для автоматизированного перевода).

Для обучения сети анализируются 3 изображения:

1. Обучающее изображение лица известного человека
2. Другая фотография того же известного человека
3. Изображение совершенно другого человека

Затем алгоритм просматривает измерения, которые он делает для каждого из этих трех изображений. Затем он немного настраивает нейронную сеть, чтобы удостовериться, что измерения, созданные для изображений #1 и #2 более похожи, а измерения для #2 и #3 – менее похожи. Схематичное представление работы алгоритма изображено на Рисунке 5.

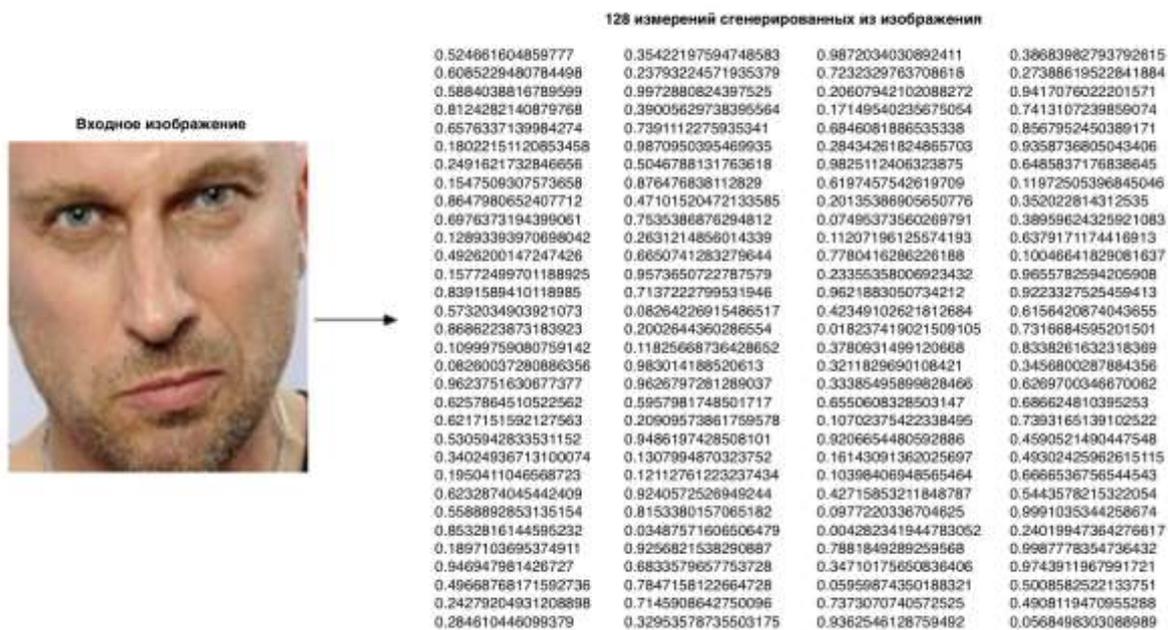


Рисунок 4 – Результаты измерений для тестового изображения



Рисунок 5 - Демонстрация работы алгоритма

Обучив сеть с различными изображениями одного и того же человека, алгоритм создает примерно одинаковые измерения. Результирующие 128 измерений называют картой лица. Идея преобразования массива необработанных данных, например, изображения, в список генерируемых компьютерных чисел крайне важна для машинного обучения (особенно для автоматизированного перевода). Используемый подход был изобретен в 2015 году в Google, он использует сверточную нейронную сеть и достиг рекордной точности 99,63% в LFW dataset.

**4 Поиск имени по карте лица:** для поиска имени человека по карте лица используется алгоритм линейного классификатора SVM (*метод опорных векторов*). В основе данного алгоритма лежит понятие плоскость решения. Плоскость разделяет полученные данные с разной классовой принадлежностью. Одним из основных преимуществ SVM можно считать высокую скорость обучения алгоритма, а, следовательно, и возможность использовать для обучения достаточно большое количество исходных данных. При помощи данного метода решаются задачи бинарной классификации, в данном примере определяется принадлежность фотографии к уже имеющейся выборке. Для этого необходимо обучить классификатор, который берет измерения из проверяемого изображения и покажет на кого больше всего похож данный человек из всей выборки обучения системы.

В качестве других классификаторов можно отметить метод релевантных векторов (RVM – Relevance Vector Machine). В отличие от SVM данный метод дает вероятности, с которыми изображение лица принадлежит данному человеку. Т.е. если SVM говорит "*изображение принадлежит человеку 1*", то RVM скажет "*изображение принадлежит человеку 1 с вероятностью  $p$  и классу человеку 2 с вероятностью  $1-p$* " [4]. Кроме того, метод опорных векторов неустойчив по отношению к шуму в исходных данных. Если обучающая выборка содержит шумовые выбросы, они существенным образом учитываются при построении разделяющей гиперплоскости. Этого недостатка лишен метод релевантных векторов.

**Заключение.** Таким образом, была реализована нейронная сеть для распознавания лиц. В рамках данной статьи были подробно описаны основные шаги, которые были необходимы для решения данной задачи. Текущая реализация алгоритма распознавания лиц является масштабируемым решением вследствие чего, полученные результаты могут быть использованы для построения сложных функциональных систем в любой области, в которой может возникнуть необходимость решения подобного рода задачи.

## ЛИТЕРАТУРА

1. Histograms of Oriented Gradients for Human Detection (In CVPR'05). N. Dalal and B. Triggs. An effective pedestrian detector based on evaluating histograms of oriented image gradients in a grid. [Электронный ресурс] // URL: <http://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf> (дата обращения: 25.12.2017).
2. V. Kazemi and S. Josephine. One millisecond face alignment with an ensemble of regression trees. In CVPR, 2014. [Электронный ресурс] // URL: <http://www.csc.kth.se/~vahidk/papers/KazemiCVPR14.pdf> (дата обращения: 25.12.2017).
3. F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In Proc. CVPR, 2015. [Электронный ресурс] // URL: [https://www.cv-foundation.org/openaccess/content\\_cvpr\\_2015/app/1A\\_089.pdf](https://www.cv-foundation.org/openaccess/content_cvpr_2015/app/1A_089.pdf) (дата обращения: 25.12.2017).
4. Christopher M. Bishop F.R.Eng. Pattern Recognition and Machine Learning. [Электронный ресурс] // URL: <https://goo.gl/WLqpHN> (дата обращения: 25.12.2017).

N.A. Popova, M.A. Nazarov, M.V. Vlasov

### **THE SOLUTION OF THE PURPOSE OF RECOGNITION OF PERSONS WITH THE USE OF MACHINE TRAINING ALGORITHMS**

*Penza State University, Penza, Russia*

*The purpose of this article is to generalize the solutions of experience and implement a neural network for face recognition. The neural network is based on special algorithms of machine learning. As an input, the algorithm receives an image with the face of one person or persons of several people, after which all persons in this image are searched using gradient histograms, the result is a fragment of the image where the basic structures of the face or persons are clearly seen. In order to determine the unique features of the face, it is necessary to take into account the difference in the angle of the face and the degree of its illumination, for this purpose, on selected fragments of the application of estimation algorithms to search for 68 points that exist on each face, it is possible to center the eyes and mouth as best as possible for more exact encoding. Encoding an image involves building an accurate "face map" consisting of 128 dimensions. Based on the search results, the convolutional neural network, using the SVM linear classifier algorithm, can determine the correspondence between different photos.*

**Keywords:** face recognition; machine learning; histogram of directed gradients; HOG; evaluation of the person's orientation; affine transformations; deep training; neural network.

## REFERENCES

1. Histograms of Oriented Gradients for Human Detection (In CVPR'05). N. Dalal and B. Triggs. An effective pedestrian detector based on evaluating histograms of oriented image gradients in a grid. [Elektronnyy resurs] // URL: <http://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf> (data obrashcheniya: 25.12.2017).
2. V. Kazemi and S. Josephine. One millisecond face alignment with an ensemble of regression trees. In CVPR, 2014. [Elektronnyy resurs] // URL: <http://www.csc.kth.se/~vahidk/papers/KazemiCVPR14.pdf> (data obrashcheniya: 25.12.2017).
3. F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In Proc. CVPR, 2015. [Elektronnyy resurs] // URL: [https://www.cv-foundation.org/openaccess/content\\_cvpr\\_2015/app/1A\\_089.pdf](https://www.cv-foundation.org/openaccess/content_cvpr_2015/app/1A_089.pdf) (data obrashcheniya: 25.12.2017).
4. Christopher M. Bishop F.R.Eng. Pattern Recognition and Machine Learning. [Elektronnyy resurs] // URL: <https://goo.gl/WLqpHN> (data obrashcheniya: 25.12.2017).